



RECEIVED

AUG 06 2002

COPY OF PAPERS  
ORIGINALLY FILED

TECH CENTER 1600/2900

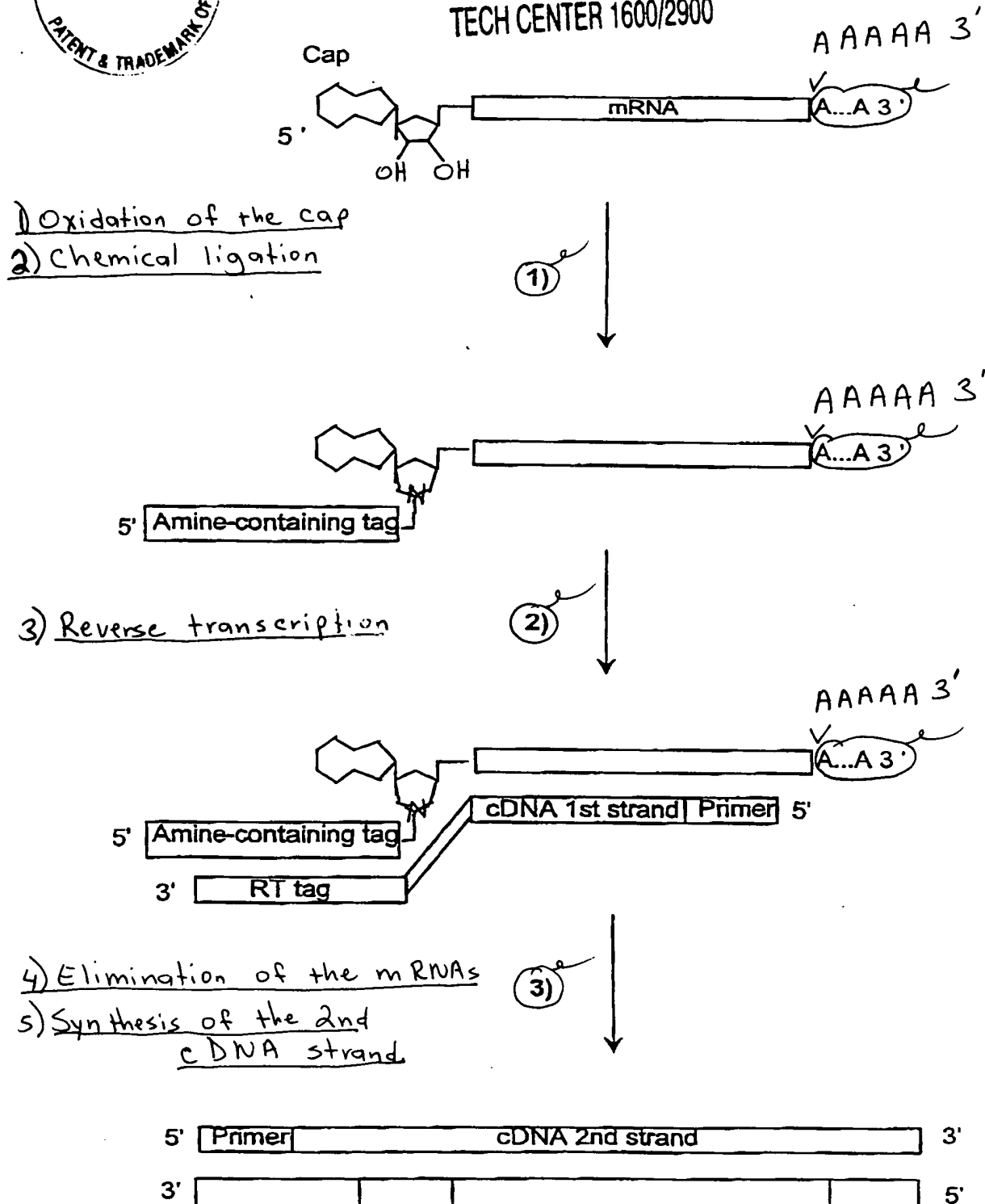


Figure 1



COPY OF PAPERS  
ORIGINALLY FILED

RECEIVED

AUG 06 2002

TECH CENTER 1600/2900

Minimum signal peptide score	false positive rate	false negative rate	proba(0.1)	proba(0.2)
3,5	0,121	0,036	0,467	0,664
4	0,096	0,06	0,519	0,708
4,5	0,078	0,079	0,565	0,745
5	0,062	0,098	0,615	0,782
5,5	0,05	0,127	0,659	0,813
6	0,04	0,163	0,694	0,836
6,5	0,033	0,202	0,725	0,855
7	0,025	0,248	0,763	0,878
7,5	0,021	0,304	0,78	0,889
8	0,015	0,368	0,816	0,909
8,5	0,012	0,418	0,836	0,92
9	0,009	0,512	0,856	0,93
9,5	0,007	0,581	0,863	0,934
10	0,006	0,679	0,835	0,919

Figure 2



COPY OF PAPERS  
ORIGINALLY FILED

RECEIVED

AUG 06 2002

TECH CENTER 1600/2900

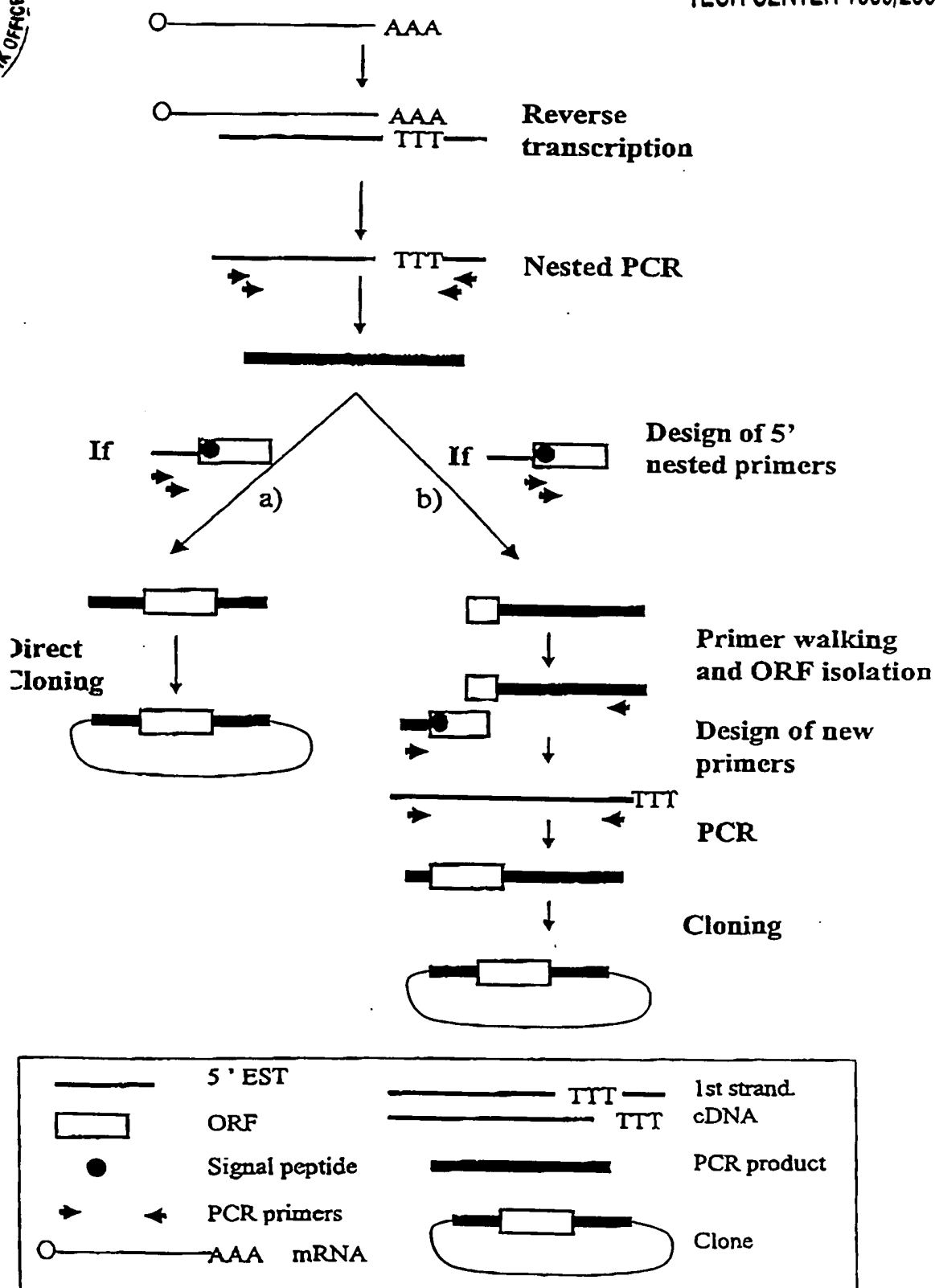


Figure 3



COPY OF PAPERS  
ORIGINALLY FILED

RECEIVED

AUG 06 2002

TECH CENTER 1600/2900

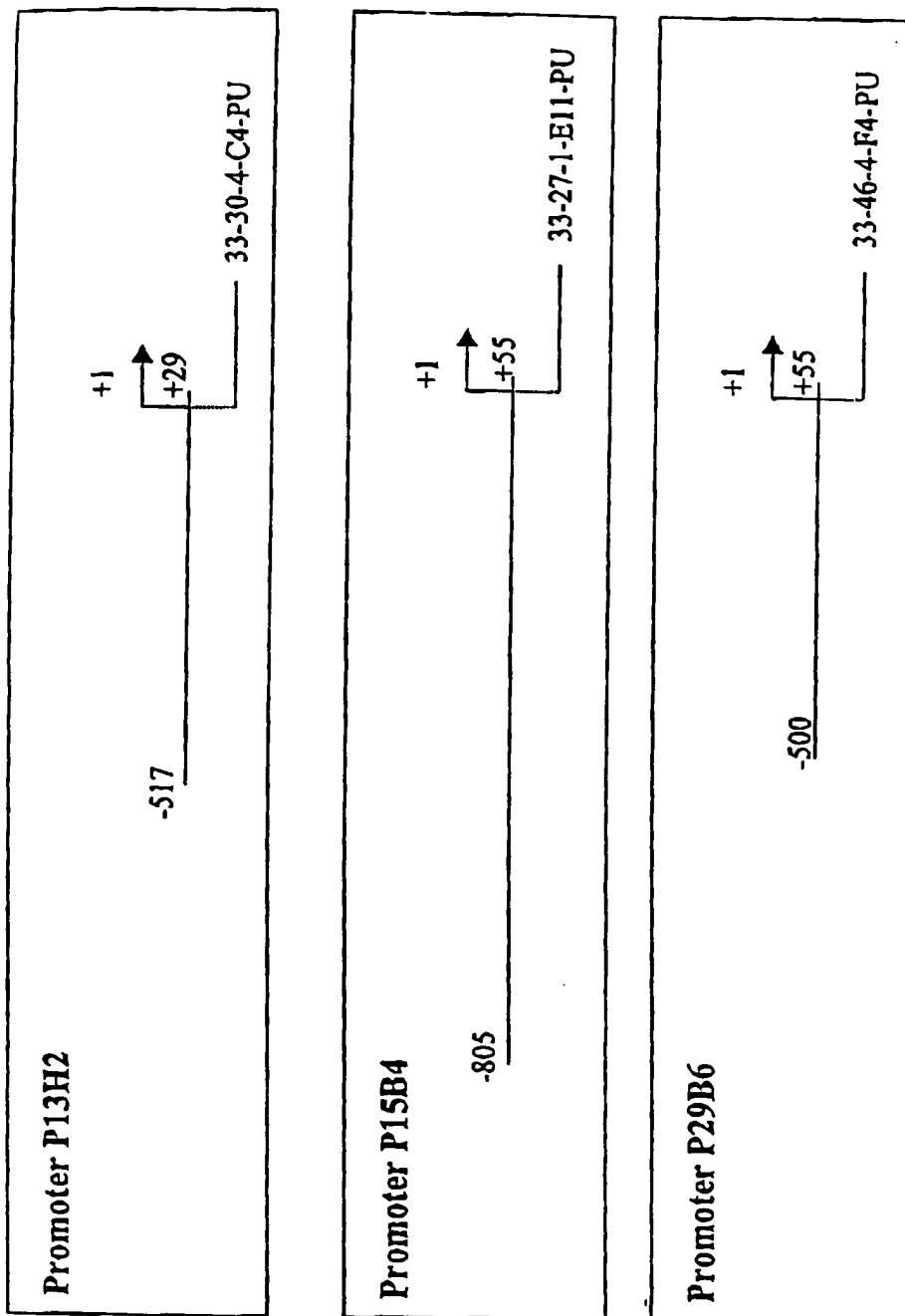


Figure 4



COPY OF PAGE  
ORIGINALLY FILED

RECEIVED

AUG 06 2002

TECH CENTER 1600/2900

Promoter sequence P13H2 (546 bp):

Matrix	Position	Orient ation	Score	Length	Sequ nc
CMYB_01	-502	+	0.983	9	TGTCAGTTG
MYOD_Q6	-501	-	0.961	10	CCCAACTGAC
S8_01	-444	-	0.960	11	AATAGAATTAG
S8_01	-425	+	0.966	11	AACTAAATTAG
DELTAEF1_01	-390	-	0.960	11	GCACACCTCAG
GATA_C	-364	-	0.964	11	AGATAAATCCA
CMYB_01	-349	+	0.958	9	CTTCAGTTG
GATA1_02	-343	+	0.959	14	TTGTAGATAGGACA
GATA_C	-339	+	0.953	11	AGATAGGACAT
TAL1ALPHAE47_01	-235	+	0.973	16	CATAACAGATGGTAAG
TAL1BETAE47_01	-235	+	0.983	16	CATAACAGATGGTAAG
TAL1BETAITF2_01	-235	+	0.978	16	CATAACAGATGGTAAG
MYOD_Q6	-232	-	0.954	10	ACCATCTGTT
GATA1_04	-217	-	0.953	13	TCAAGATAAAGTA
IK1_01	-126	+	0.963	13	AGTTGGGAATTCC
IK2_01	-126	+	0.985	12	AGTTGGGAATTCC
CREL_01	-123	+	0.962	10	TGGGAATTCC
GATA1_02	-96	+	0.950	14	TCAGTGATATGGCA
SRY_02	-41	-	0.951	12	TAAACAAAACA
E2F_02	-33	+	0.957	8	TTTAGCGC
MZF1_01	-5	-	0.975	8	TGAGGGGA

Promoter sequence P15B4 (861bp) :

Matrix	Position	Orient ation	Score	Length	Sequence
NFY_Q6	-748	-	0.956	11	GGACCAATCAT
MZF1_01	-738	+	0.962	8	CCTGGGGA
CMYB_01	-684	+	0.994	9	TGACCGTTG
VMYB_02	-652	-	0.985	9	TCCAACGGT
STAT_01	-673	+	0.968	9	TTCCTGGAA
STAT_01	-673	-	0.951	9	TTCCAGGAA
MZF1_01	-556	-	0.956	8	TTGGGGGA
IK2_01	-451	+	0.965	12	GAATGGGATTTC
MZF1_01	-424	+	0.986	8	AGAGGGGA
SRY_02	-398	-	0.955	12	GAAAACAAAACA
MZF1_01	-216	+	0.960	8	GAAGGGGA
MYOD_Q6	-190	+	0.981	10	AGCATCTGCC
DELTAEF1_01	-176	+	0.958	11	TCCCACCTTCC
S8_01	5	-	0.992	11	GAGGCAATTAT
MZF1_01	16	-	0.986	8	AGAGGGGA

Promoter sequence P29B6 (555 bp) :

Matrix	Position	Orient ation	Score	Length	Sequence
ARNT_01	-311	+	0.964	16	GGACTCACGTGCTGCT
NMYC_01	-309	+	0.965	12	ACTCACGTGCTG
USF_01	-309	+	0.985	12	ACTCACGTGCTG
USF_01	-309	-	0.985	12	CAGCACGTGAGT
NMYC_01	-309	-	0.956	12	CAGCACGTGAGT
MYCMAX_02	-309	-	0.972	12	CAGCACGTGAGT
USF_C	-307	+	0.997	8	TCACGTGC
USF_C	-307	-	0.991	8	GCACGTGA
MZF1_01	-292	-	0.968	8	CATGGGGA
ELK1_02	-105	+	0.963	14	CTCTCCGGAAGCCT
CETS1P54_01	-102	+	0.974	10	TCCGGAAGCC
AP1_Q4	-42	-	0.963	11	AGTGACTGAAC
AP1FJ_Q2	-42	-	0.961	11	AGTGACTGAAC
PADS_C	45	+	1.000	9	TGTGGTCTC

Figure 5 — Deleted



RECEIVED

AUG 06 2002

TECH CENTER 1600/2900

COPY OF PAPERS  
ORIGINALLY FILED

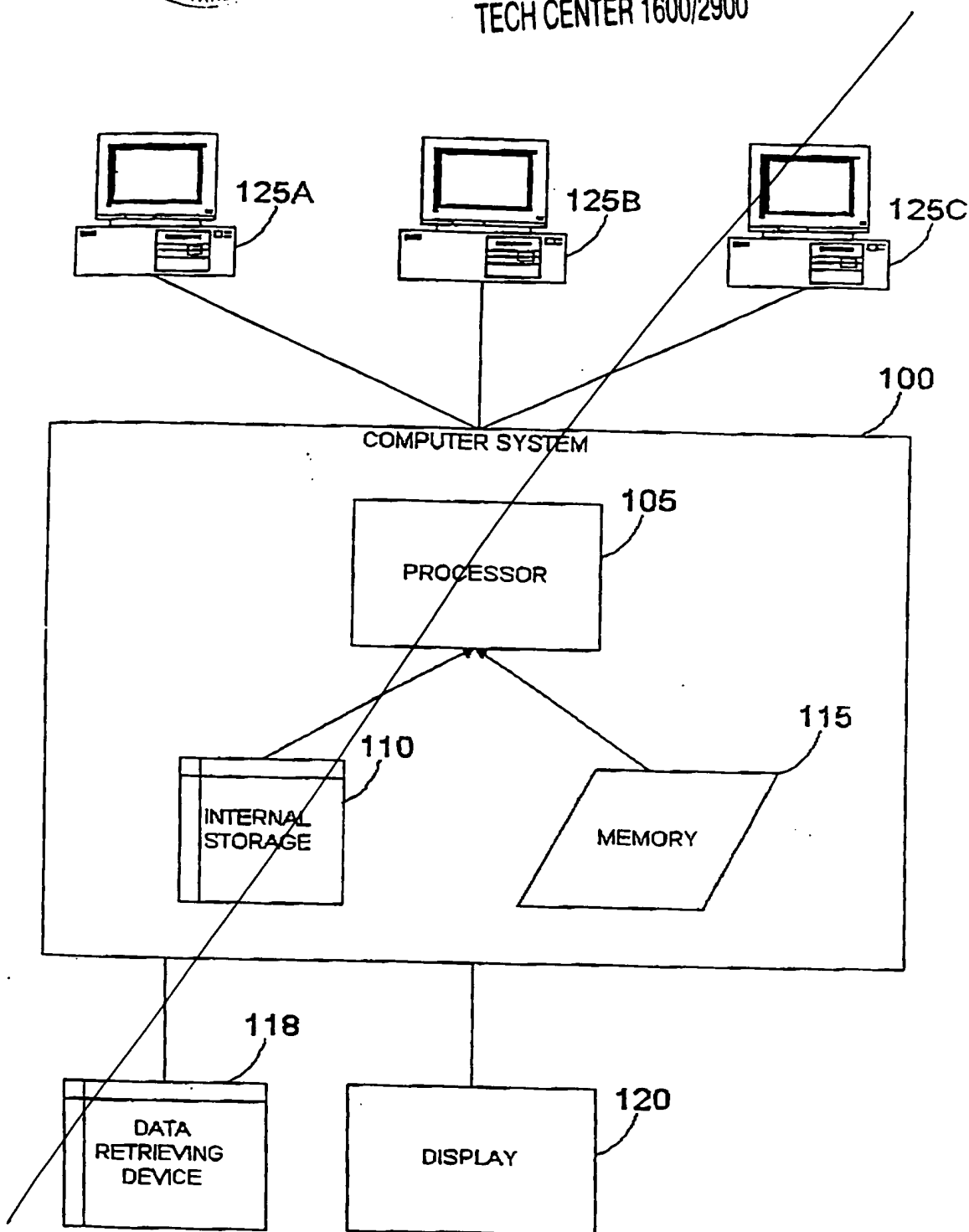


FIGURE 6 - Deleted



COPY OF PAPERS  
ORIGINALLY FILED

RECEIVED

AUG 06 2002

TECH CENTER 1600/2900

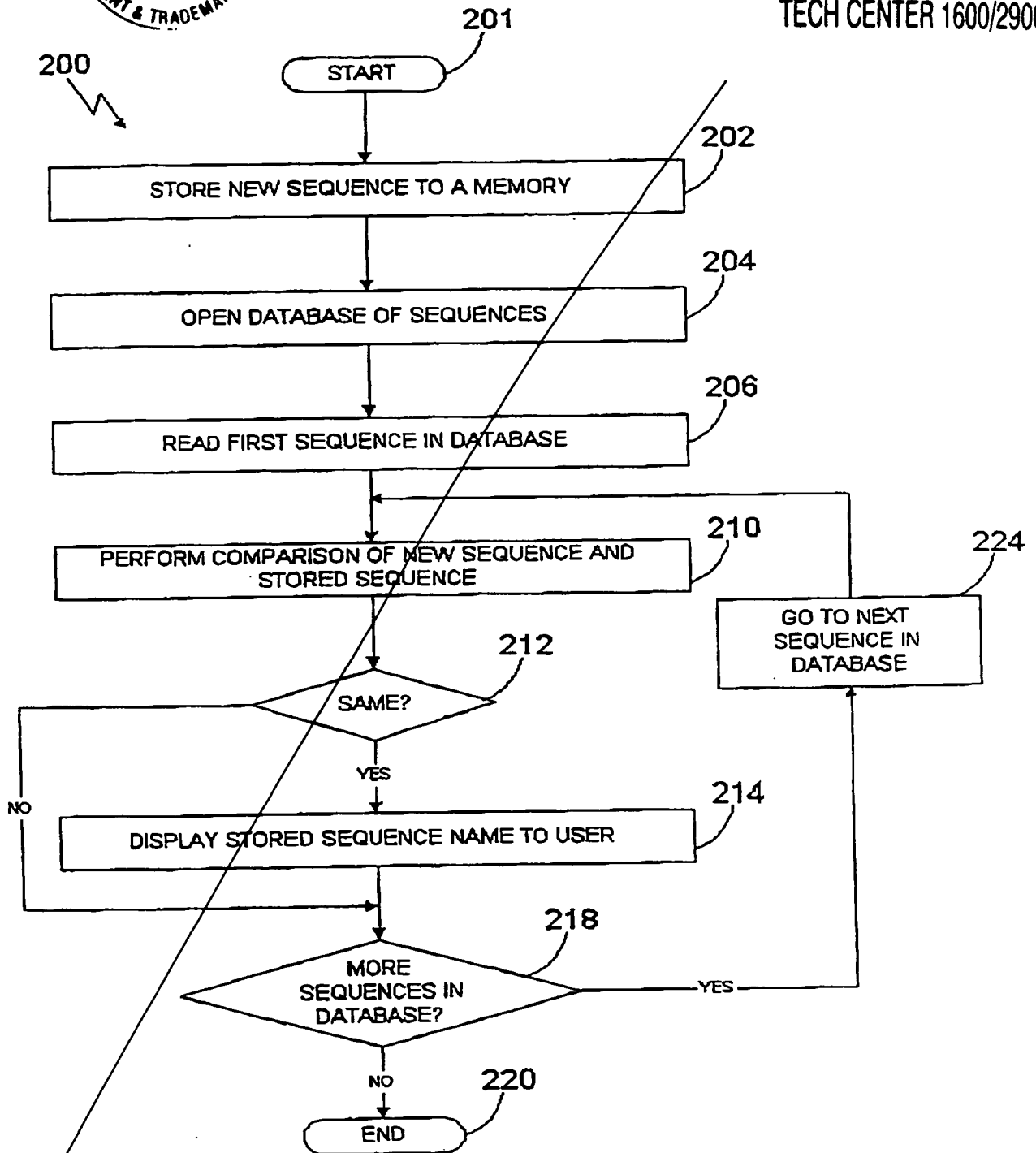


FIGURE 7 - Deleted



COPY OF PAPERS  
ORIGINALLY FILED

RECEIVED

AUG 06 2002

TECH CENTER 1600/2900

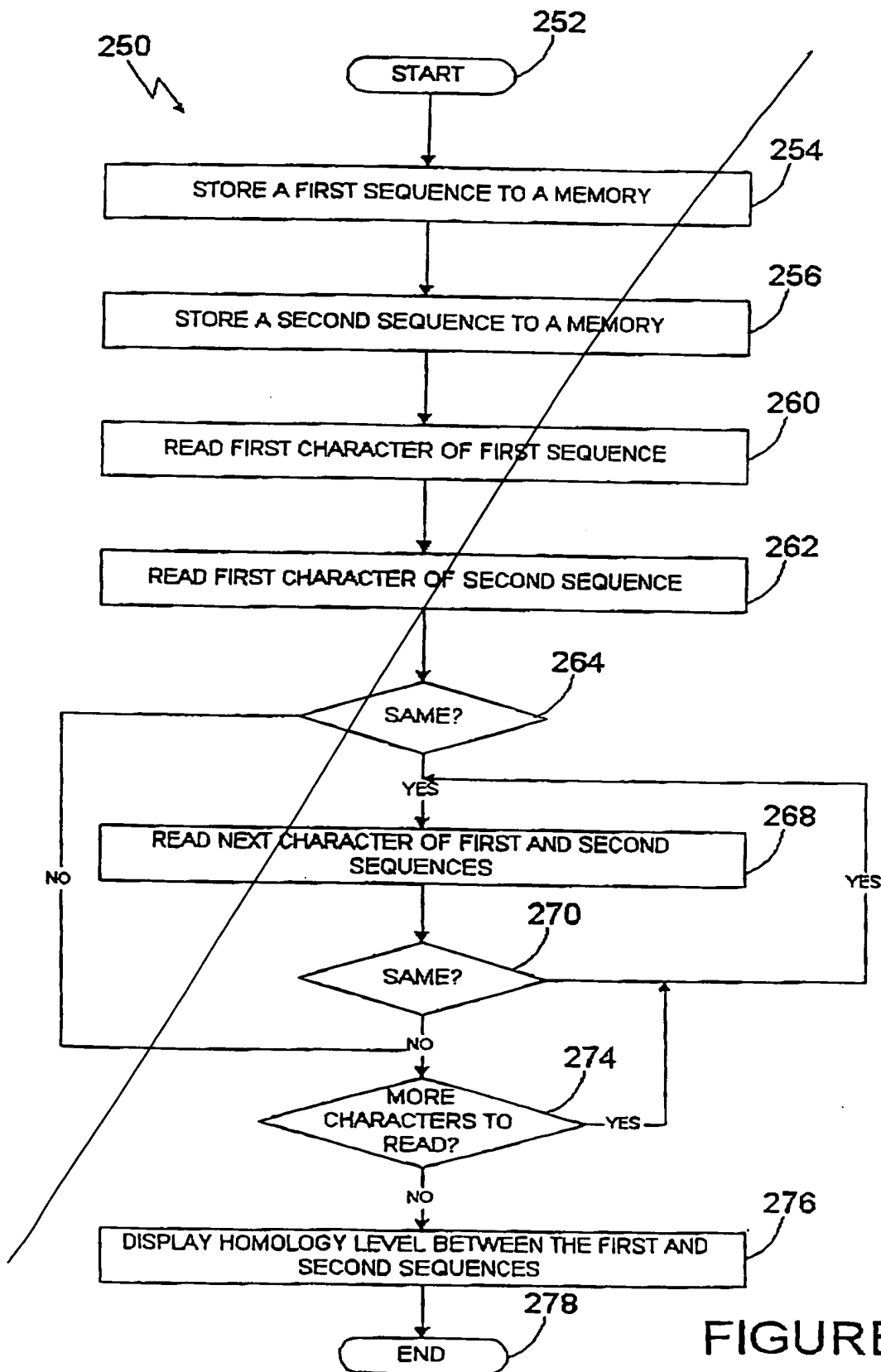


FIGURE 8 -  
Deleted



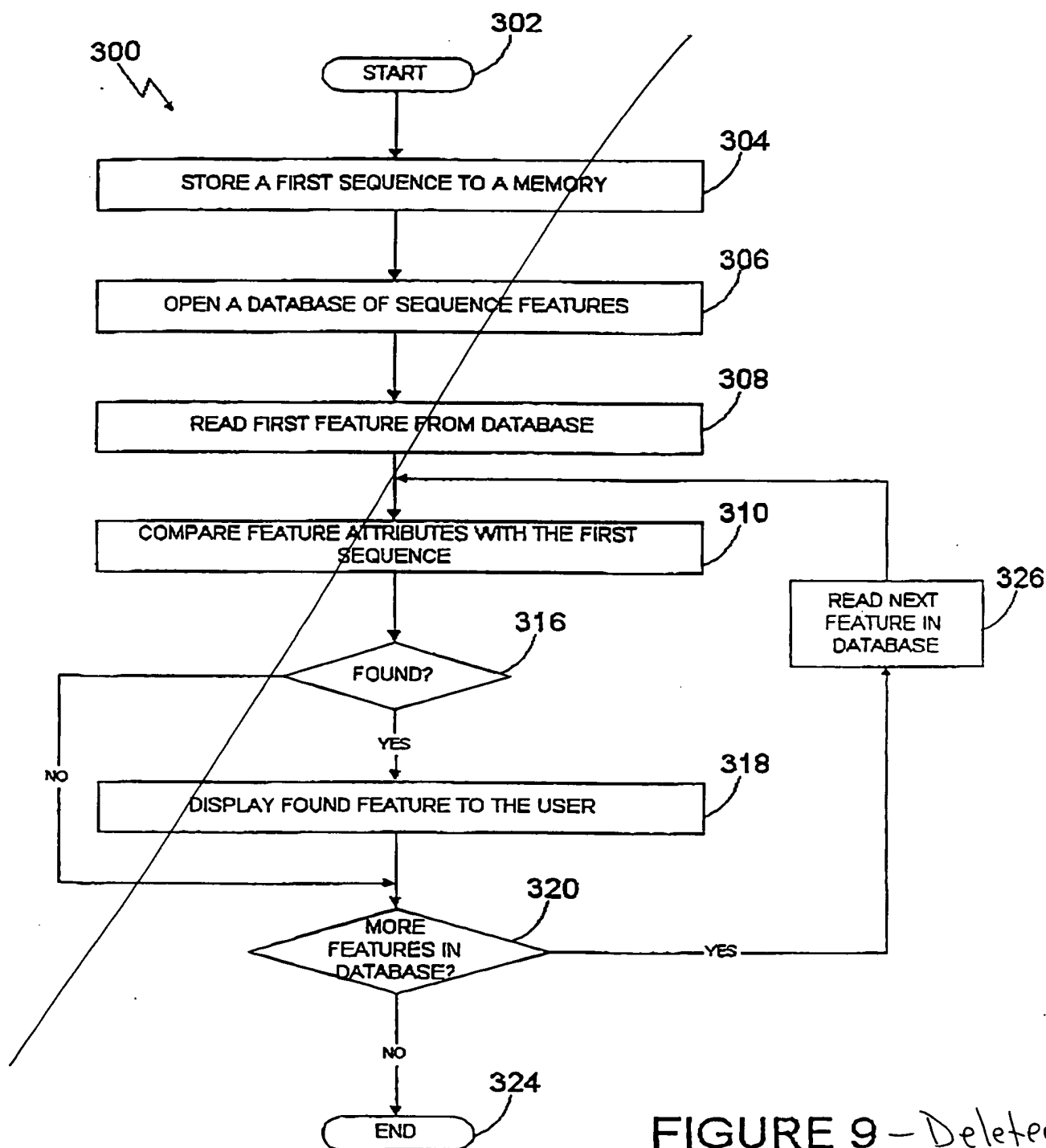


COPY OF PAPERS  
ORIGINALLY FILED

RECEIVED

AUG 06 2002

TECH CENTER 1600/2900





COPY OF PATENT  
ORIGINALLY FILED

RECEIVED

AUG 06 2002

TECH CENTER 1600/2900

Search characteristic		Selection Characteristics			Comments
Step	Program	Strand	Parameters	Identity (%)	
miscellaneous	FASTA	both	-	90	15
tRNA	FASTA	both	-	80	60
rRNA	BLASTN	both	S=108	80	40
mtRNA	BLASTN	both	S=108	80	40
Prokaryotic	BLASTN	both	S=144	90	40
Fungal	BLASTN	both	S=144	90	40
Alu	BLASTN	both	S=72, B=5	70	40
L1	BLASTN	both	S=72, B=5	70	40
Repeats	BLASTN	both	S=72	70	40
PolyA	BLAST2N	top	W=6, S=10, E=1000, N=12	90	10
Polyadenylation signal	-	top	AAATAAA allowing 1 mismatch		
Vertebrate	BLASTN then FASTA	both	-	90 then 70	30
ESTs	BLAST2N	both	-	90	30
Geneseq	BLASTN	both	W=8, B=10	90	30
ORF	BLASTP	top	W=8, B=10	-	-
Proteins	BLASTX	top	E = 0.001	70	30

Figure 10

Fig. 5

in the 50 nucleotides before the 5' end of the polyA  
first BLASTN, then FASTA on matching sequences

in the last 100 nucleotides

on ORF proteins, max 10 matches



COPY OF PAPERS  
ORIGINALLY FILED

RECEIVED

AUG 06 2002

TECH CENTER 1600/2900

1/5

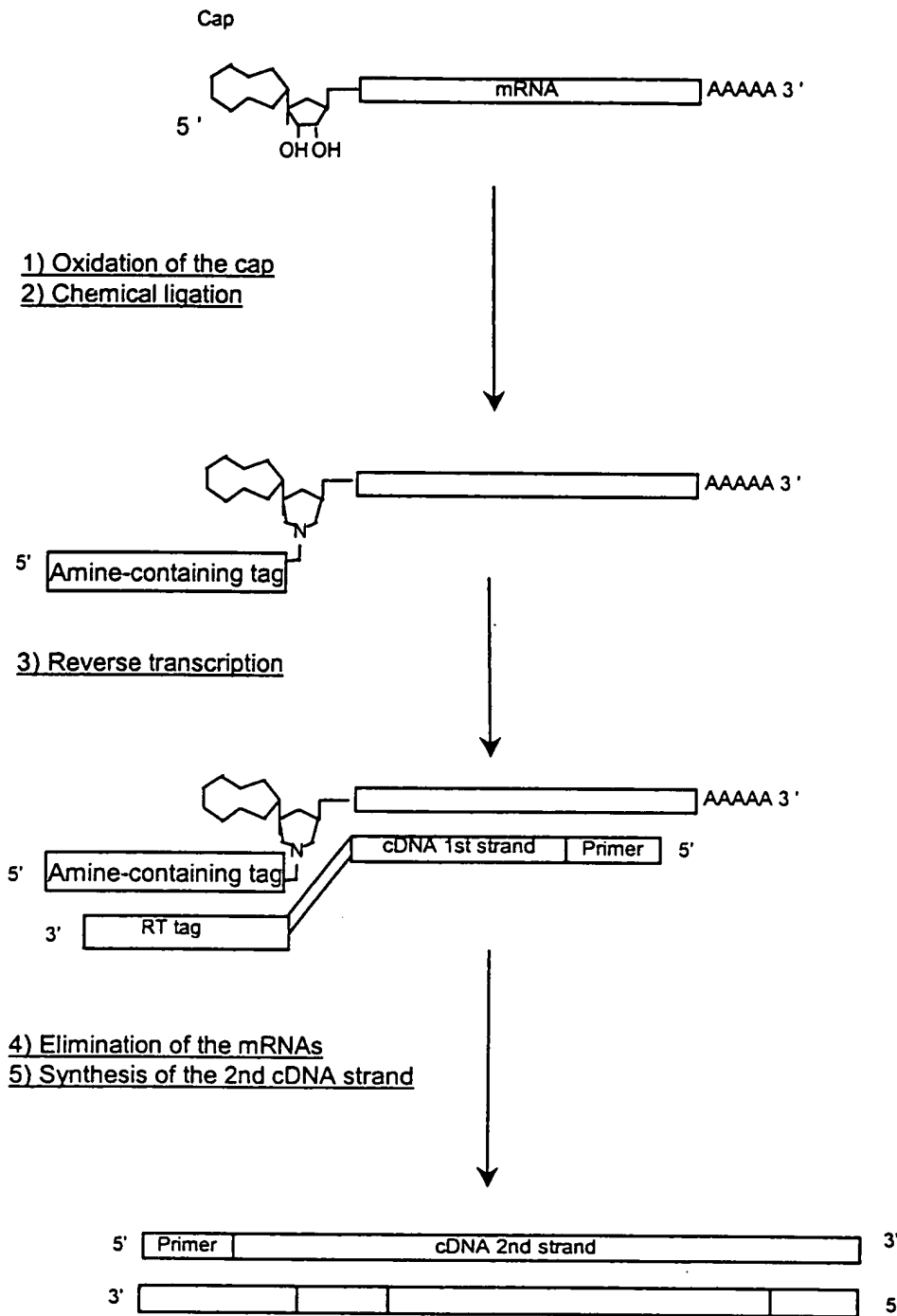


Figure 1



COPY OF PAPERS  
ORIGINALLY FILED

RECEIVED

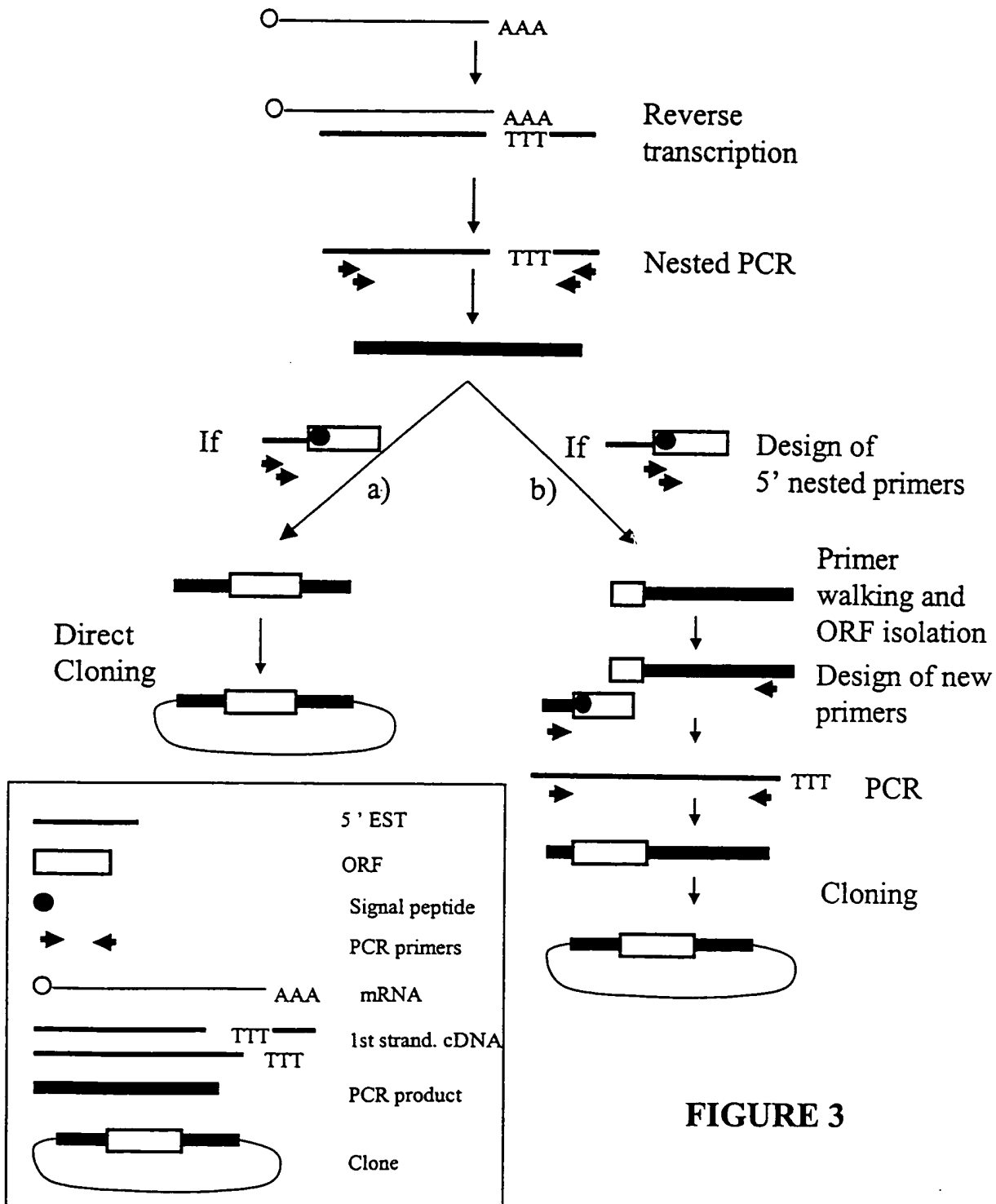
AUG 06 2002

TECH CENTER 1600/2900

2/5

Minimum signal peptide score	false positive rate	false negative rate	proba(0.1)	proba(0.2)
3.5	0.121	0.036	0.467	0.664
4	0.096	0.06	0.519	0.708
4.5	0.078	0.079	0.565	0.745
5	0.062	0.098	0.615	0.782
5.5	0.05	0.127	0.659	0.813
6	0.04	0.163	0.694	0.836
6.5	0.033	0.202	0.725	0.855
7	0.025	0.248	0.763	0.878
7.5	0.021	0.304	0.78	0.889
8	0.015	0.368	0.816	0.909
8.5	0.012	0.418	0.836	0.92
9	0.009	0.512	0.856	0.93
9.5	0.007	0.581	0.863	0.934
10	0.006	0.679	0.835	0.919

Figure 2





COPY OF PAGE  
ORIGINALLY FILED

RECEIVED

AUG 06 2002

TECH CENTER 1600/2900

4/5

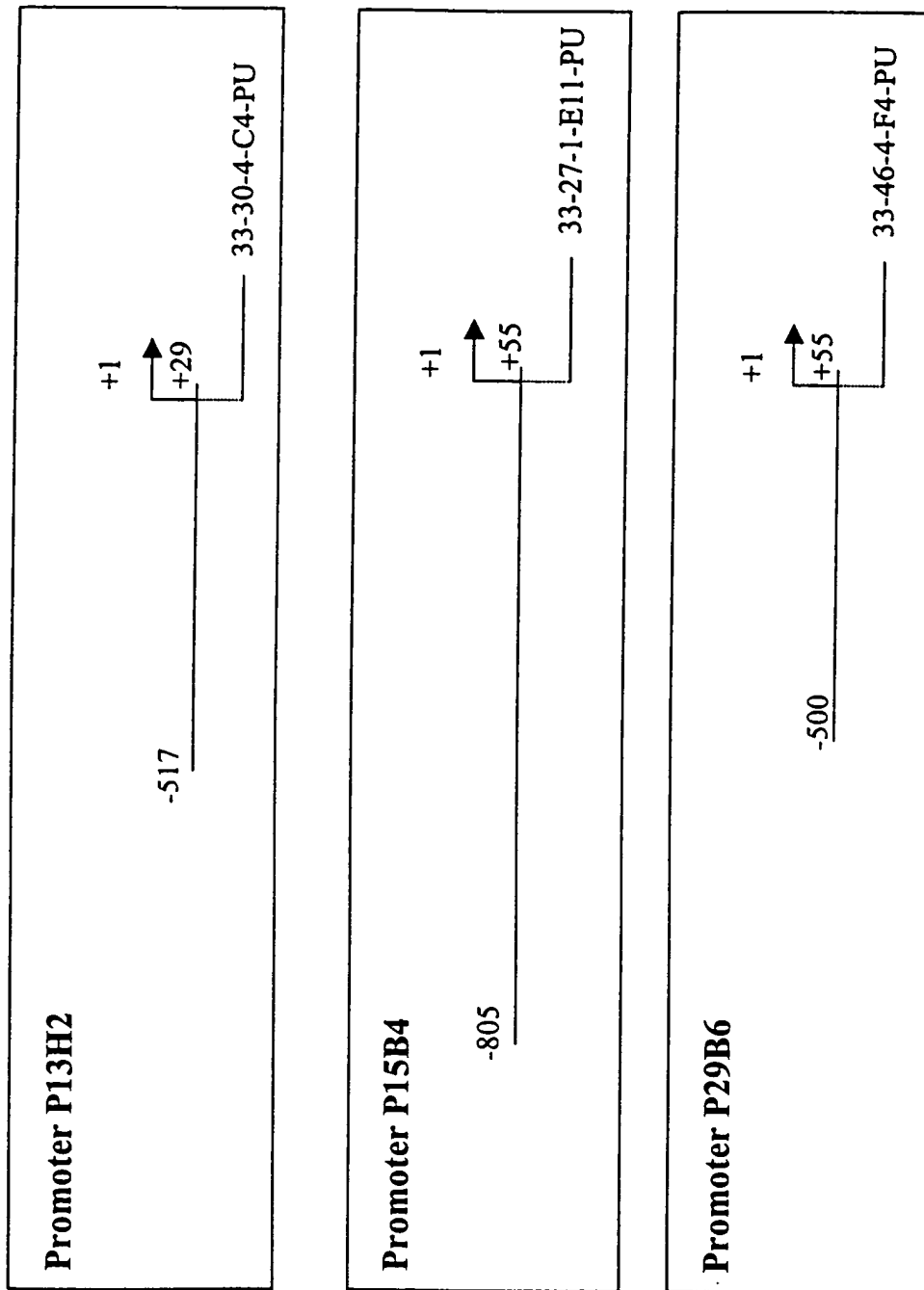


Figure 4



COPY OF PAPERS  
ORIGINALLY FILED

RECEIVED

AUG 06 2002

5/5

TECH CENTER 1600/2900

Step	Search characteristic		Selection Characteristics			
	Program	Strand	Parameters	Identity (%)	Length (bp)	Comments
miscellaneous	FASTA	both	-	90	15	
tRNA	FASTA	both	-	80	60	
rRNA	BLASTN	both	S=108	80	40	
mtRNA	BLASTN	both	S=108	80	40	
Prokaryotic	BLASTN	both	S=144	90	40	
Fungal	BLASTN	both	S=144	90	40	
Alu	BLASTN	both	S=72, B=5	70	40	max 5 matches, masking
L1	BLASTN	both	S=72, B=5	70	40	max 5 matches, masking
Repeats	BLASTN	both	S=72	70	40	masking
PolyA	BLAST2N	top	W=6, S=10, E=1000, N=-12	90	10	in the last 100 nucleotides
Polyadenylation signal	-	top	AATAAA allowing 1 mismatch			in the 50 nucleotides preceding the 5' end of the polyA
Vertebrate	BLASTN then FASTA	both	-	90 then 70	30	first BLASTN and then FASTA on matching sequences
ESTs	BLAST2N	both	-	90	30	
Geneseq	BLASTN	both	W=8, B=10	90	30	
ORF	BLASTP	top	W=8, B=10	-	-	on ORF proteins, max 10 matches
Proteins	BLASTX	top	E = 0.001	70	30	

Parameters used for each step of cDNA analysis

Figure 5